Knowledge-Based Systems 37 (2013) 186–195

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon

Fu Xianghua *, Liu Guo, Guo Yanyan, Wang Zhiqiang

College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China

ARTICLE INFO

Article history:
Received 17 October 2011
Received in revised form 25 July 2012
Accepted 3 August 2012
Available online 27 September 2012

Keywords:
Aspect detection
Sentiment analysis
Social reviews
Topic modeling
HowNet lexicon

ABSTRACT

User-generated reviews on the Web reflect users' sentiment about products, services and social events. Existing researches mostly focus on the sentiment classification of the product and service reviews in document level. Reviews of social events such as economic and political activities, which are called social reviews, have specific characteristics different to the reviews of products and services. In this paper, we propose an unsupervised approach to automatically discover the aspects discussed in Chinese social reviews and also the sentiments expressed in different aspects. The approach is called Multi-aspect Sentiment Analysis for Chinese Online Social Reviews (MSA-COSRs). We first apply the Latent Dirichlet Allocation (LDA) model to discover multi-aspect global topics of social reviews, and then extract the local topic and associated sentiment based on a sliding window context over the review text. The aspect of the local topic is identified by a trained LDA model, and the polarity of the associated sentiment is classified by HowNet lexicon. The experiment results show that MSA-COSR cannot only obtain good topic partitioning results, but also help to improve sentiment analysis accuracy. It helps to simultaneously discover multi-aspect fine-grained topics and associated sentiment.

1. Introduction

Social media is a group of Internet-based applications that are built on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content [25]. Compared with traditional media, social media introduces substantial and pervasive changes to communication between organizations, communities, and individuals [2]. The rise of the social media motivates people to freely express their sentiment and opinions about anything more frequently than ever before. For most commercial organizations and government departments, the online reviews represent invaluable source of information. There are many special teams dedicated to the task of reading the content posted on social media sites and extracting public opinions expressed on their products, services and policies. However, with the dramatic growth of online reviews, it is becoming increasingly difficult to analyze online reviews by hand and to take immediate, real-time action. Hence, in recent years, there have been many interests in the natural language processing and data mining communities to develop text mining techniques with the capability of accurately extracting people's opinions from large volumes of unstructured review text [21].

Sentiment analysis or opinion mining aims to automatically detect subjective information such as opinions, attitudes, and feelings expressed in text. Much work has been done to extract information from reviews, summarize user's opinions, and categorize reviews according to opinion polarities [10,13,21,27]. The review mining tasks have been studied ranging from coarse-grained document level polarity classification [22] to fine-grained extraction of opinion expressions and their targets [18,37]. Nevertheless, with the existing techniques, it is still hard for users to easily digest and exploit the large number of reviews due to inadequate support for understanding individual reviewer's opinions at the fine-grained level of topical aspects [35]. For example, hotel reviews usually discuss multiple aspects of environment, food, charge, service, etc. Even though some reviewers give identical overall ratings, their feelings on various aspects can be different. So it is necessary to achieve deeper and more detailed understanding of the reviews, and analyze the underlying sentiment on each aspect.

Most research works on review mining pay attentions only to product reviews [18,21,27]. There are few works focusing on the social reviews mining [8,16,19,20]. Recent works on topic sentiment modeling also fail to distinguish product reviews and social reviews [13,30]. In this paper, we dedicate to analyze sentiment in online social reviews. Product reviews generally emerge on electronic business sites. The evaluation targets of product reviews usually compose of specific entities, the syntax structure of the product reviews generally is not complex, and there are ordinarily

* Corresponding author.
E-mail address: fuxh@szu.edu.cn (F. Xianghua).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.08.003

a whole rating associated with the reviews. Social reviews are different from product reviews in that they usually are located on Bulletin Board System (BBS), Blog, and micro-blogging sites. Social reviews are usually on hot economic and political events of far-ranging background knowledge and uncertain evaluation targets. The reviews of economic and political events cover wide range topics, encompassing a variety of attitude judgments toward various entities and issues. The syntax structure and grammar, such as metaphor and irony, could be more complex in the social reviews. These attitude judgments often interact in unexpected and counterintuitive ways. Furthermore, there are no additional ratings associated with the social reviews, which makes it difficult to get labeled dataset and apply supervised learning methods. Compared with product review mining, analyzing sentiment of the social reviews is more challenging task.

In this paper, we propose unsupervised approach to determine the multi-aspect sentiment of Chinese social reviews, which is called Multi-aspect Sentiment Analysis for Chinese Online Social Reviews (MSA-COSRs). We first apply the Latent Dirichlet Allocation (LDA) model to discover multi-aspect global topics of the social reviews set, and then extract the local topic and associated sentiment based on a sliding window context over the review text. We also give the algorithms about how to identify the local topic and classify the sentiment orientation. MSA-COSR can help to simultaneously identify multi-aspect topics and the associated sentiment of these topics. The experiment results show that MSA-COSR cannot only obtain good topic partitioning results, but also help to improve the accuracy of sentiment analysis.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works. The principle of local topics discovery based on LDA topic model is proposed in Section 3. Section 4 introduces a method to analyze sentiment of a sliding window based on HowNet lexicon. Experiments and evaluations are reported in Section 5. We conclude the paper in Section 6 with future researches.

2. Related works

In recent years, there have been many research interests in sentiment analysis and opinion mining on review text. Sentiment analysis methods on text usually can be divided into two categories: (1) the first type is based on part-of-speech (POS) tagging of words and sentiment lexicons. This type is proposed by Peter [32] at first. Both Xu et al. [38] and Du et al. [4] adopted this type of method for the sentiment analysis of Chinese online reviews. Liu and Li [15] proposed to establish a believable vocabulary on semantic knowledge named HowNet, and then get the sentiment polarity of words through comparison with the similarity between the words. Zhu [43] succeeded in judging semantic orientation of Chinese online reviews based on the HowNet lexicon. (2) The second type is based on machine learning algorithms. For example, Whitelaw used support vector machine (SVM) to classify the sentiment orientation of movie reviews [36]. Lin proposed a learning method based on statistical model, which obtained opinions reflected in the text through analyzing words [14]. Some other researchers also used machine learning algorithms to classify the sentiment orientation of Chinese online reviews, e.g., Tang et al. [28] adopted several supervised machine learning methods to classify the sentiment of Chinese reviews, Liao et al. made use of the probability statistical model to retrieval blog opinions [12], Xu et al. used Naive Bayesian (NB) and maximum entropy [39] to classify sentiment of Chinese news, Hu et al. compared the qualities of sentiment classification between supporting vector machine and Naïve Bayesian classifier [7]. Many works have been focused on the problem of sentiment classification at various levels using machine learning techniques. Turney and Littman [33] applied an unsupervised learning algorithm to classify the semantic orientation in the word/phrase level, based on mutual information between document phrases and a small set of positive/negative paradigm words like ‘‘good” and ‘‘bad”.

Furthermore, the sentiment classification tasks also were processed in different language grains. Such as, Turney and Littman [34] classified the semantic orientation in the word/phrase level, based on mutual information between document phrases and a small set of positive/negative paradigm words like ‘‘good” and ‘‘bad”; Ye et al. [41] applied supervised learning method to classify the sentiment orientation of trval reviews in document level; and Ganapathibhotla et al. [5] proposed a technique to ming opinions in comparative sentences; and Su et al. [27] proposed an co-clustering algorithm to Ming the opinions of products' features.

Although researchers have got many achievements, the existing researches mostly focused on the sentiment classification of the product and service reviews in document level. Only a few researchers pay attention to the sentiment analysis of the social reviews. There are some research works on identifying the overall sentiment polarities to political reviews. For example, Mullen and Malouf [19] described preliminary statistical tests on a new dataset of political discussion group postings, which indicate that posts made in direct response to other posts in a thread have a strong tendency to represent an opposing political viewpoint to the original post. They conclude that traditional text classification methods are inadequate to handle the task of sentiment analysis in this domain. Malouf and Mullen [16] used a variety of sentiment analysis, text classification, and social network analysis methods to evaluate against the user's self-descriptions. They found purely text-based methods performed poorly. But they could be improved by using techniques taking into account the user' position in the online community. Somasundaran and Wiebe [26] explored the utility of sentiment and arguing opinions for classifying stances in ideological debates. In order to capture arguing opinions in ideological stance taking, they constructed an arguing lexicon automatically from a manually annotated corpus. They found, by employing sentiment and arguing opinions and their targets as features of supervised system, it is possible to perform better than a unigram based system. Mei et al. [17] proposed Topic-Sentiment Mixture (TSM) model to reveal the latent topical facets in a Weblog collection. With a specifically designed HMM structure, the sentiment models and topic model estimated with TSM can be utilized to extract topic life cycles and sentiment dynamics. O'Hare et al. [20] found that there exists topic shift within financial Weblogs. To deal with the problem of topic shift, they proposed text extraction techniques to create topic-specific sub-documents, which were used to train the sentiment classifier. In Chinese social reviews' sentiment analysis, Tao et al. [29] proposed an approach for feature extraction of sentiment analysis of the news comments, which can provide finer-grained sentiment analysis for specific news topic. The candidate sentimental features were extracted according to the comparison between contents of the news comments and the corresponding news. Then the general sentiment features were selected by several extension and validation processes. Yang et al. [40] attempted to construct a new sentiment lexicon with sentiment orientation extent based on existing HowNet and NTUSD, which was applied in a semi-automatic Web public opinion analysis system.

Recently, topic models have been introduced for simultaneous analysis of topics and sentiment in a document. These studies, which jointly model topic and sentiment, take the advantage of the relationship between topics and sentiment, and are shown to be superior to traditional sentiment analysis tools. One of the most closely related works is the Topic-Sentiment Model (TSM) [17], which jointly modeled the mixture of topics and sentiment predic-

tions for the entire document. Because TSM is essentially based on the Probabilistic Latent Semantic Indexing (pLSI) model with an extra background component and two additional sentiment subtopics, it suffers from the problems of inference on new document and overfitting the data, both of which are known as the deficits of pLSI. Titov and McDonal [31] proposed the Multi-grain Latent Dirichlet Allocation model (MG-LDA), which is also closely related to our work, since they are all based on the state-of-the-art topic model LDA. MG-LDA is argued to be more appropriate to build topics that are representative of ratable aspects of objects from either a global topic or a local topic. Being aware of the limitation that MG-LDA is still purely topic based without considering the associations between topics and sentiments, Titov and McDonald [30] further proposed the Multi-Aspect Sentiment Model (MAS) by extending the MG-LDA framework. The major improvement of MAS is that it can aggregate sentiment texts for the sentiment summary of each rating aspect extracted from the MG-LDA. But MAS works on a supervised setting as it requires that every aspect is rated at least in some documents, which is practically infeasible in real-world applications. Lin and He [13] proposed a probabilistic modeling framework called Joint Sentiment/Topic Model (JST), which detects sentiment and topic simultaneously from text. Unlike other machine learning approaches to sentiment classification which often require labeled corpus for classifier training, the JST model is fully unsupervised. But JST only predicts the sentiment orientation in the document level.

Some other related topic-sentiment modeling works include: Li et al. [11] proposed the Sentiment-LDA model and Dependency-Sentiment-LDA. Unlike the previous models making assumption that the sentiments of the words in the document are all independent, this model views the sentiments of words as a Markov chain. Zhao et al. [42] proposed a MaxEnt-LDA hybrid model to jointly discover both aspects and aspect-specific opinion words. They show that with a relatively small amount of training data, MaxEnt-LDA hybrid model can effectively identify aspect and opinion words simultaneously.

To the best of our knowledge, there is still no work on multi-aspect sentiment analysis of the Chinese online social reviews. We propose MSA-COSR to evaluate the multi-aspect sentiment of Chinese social reviews in this paper. Compared with other methods, this approach is different in the following aspects: (1) MSA-COSR is the first to evaluate multi-aspect sentiment of the Chinese social review in topic level; (2) MSA-COSR is an unsupervised method, which does not need labeled training dataset that is the characteristic of social reviews; and (3) Because LDA topic model and HowNet lexicon are relatively proven techniques, the implementation of MSA-COSR is not difficult.

Because the conditions and targets of topic-sentiment models are different, it is difficult to compare the aspects and sentiments found by these models. We evaluate our MSA-COSR with a Chinese online social review dataset in document level and aspect level, and compare the result of sentiment identification with some supervised methods.

3. Multi-aspect local topic discovery based on LDA topic model

A reviewer usually expresses different opinions about several topics in a social review text at the same time, so it is insufficient to merely evaluate the sentiment in document levels or in sentence levels. To deeply understand the reviewer's feeling, one important subtask is to uncover the multi-aspect latent topics the reviewer is discussing, and another important subtask is to discover how opinions and sentiments of different aspects are expressed. We train a LDA model to discover global topics, and then identify local topics and sentiments with a sliding window method.

3.1. Introduction to LDA topic model

Probabilistic graphical models have gained tremendous attention in machine learning and nature language processing research areas in recent years. LDA model [1] is one of the most popular topic models, which is based on the assumption that each document is a mixture of various topics, and each topic is a probability distribution over different words.

Assume that we have a corpus with a collection of documents denoted by $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$, the corresponding vocabulary of the document collection denoted by $\mathcal{V} = \{w_1, w_2, \ldots, w_{|\mathcal{V}|}\}$. Each document in the corpus is a sequence with dN words denoted by $\mathbf{d} = \{w_{d1}, w_{d2}, \ldots, w_{dN}\}$, where $|\mathcal{D}|$ is the document number of the collection $\mathcal{D}$, $|\mathcal{V}|$ is the number of words, and w_{dn} is the nth word in the document $\mathbf{d}$, $w_{dn} \in \mathcal{V}$. Also we assume that the latent topic set discussed in the document collection $\mathcal{D}$ is denoted by $\mathcal{K} = \{\phi_1, \phi_2, \ldots, \phi_{|\mathcal{K}|}\}$, where $|\mathcal{K}|$ is the topic number of $\mathcal{K}$. LDA model takes into account that each document is a probabilistic mixture of the topic set $\mathcal{K}$, and each topic is a multinomial distribution of the words in vocabulary $\mathcal{V}$. Fig. 1 shows the graphical model of LDA model.

In LDA model, generation of a collection is started by sampling word distributions $\phi_{1:|\mathcal{K}|}$ from a prior Dirichlet distribution $Dir(\beta)$ for each latent topic. Then each document d is generated as follows:

(1) choose the distribution of topics $\theta_d \sim Dir(\alpha)$,
(2) for each word w_{dn} in the document $\boldsymbol{d}$,
  – Choose topic $z_{d,n} \sim \theta_d$.
  – Choose word $w_{dn} \sim \phi_{z_{d,n}}$.

LDA model has a distinct hierarchical structure, which includes the document collection layer, the document layer, and the word layer. LDA model can be determined by two parameters α and β of document collection layer, where α is a prior super parameters about topic distribution in the document collection, which means the generation probability of among topics; and β characters the word distribution of each latent topic, which can parameterize by a $|\mathcal{K}| \times |\mathcal{V}|$ matrix ($\beta_{ij} = p(w_j = 1|z_i = 1)$), and the kth row of the matrix represents the distribution of the kth topic ϕ_k. In the document layer, θ_d is the distribution of all the topics in document d, which obeys a Dirichlet distribution $Dir(\alpha)$. In the word layer, z_{dn} is a distribution of the topics which generates the word w_{dn}, $z_{dn} \sim multinominal(\theta_d)$.

The exact inference of learning parameters in LDA is intractable due to the coupling between the topics z and β [1]. Therefore approximate techniques are needed. The most common approaches that are used for approximate inference are Gibbs Sampling and variational method [1,6]. After the model parameters have been determined, given a document d, the posterior probability of the document d about the latent topic θ is defined as:

$$p(\theta, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)} \tag{1}$$

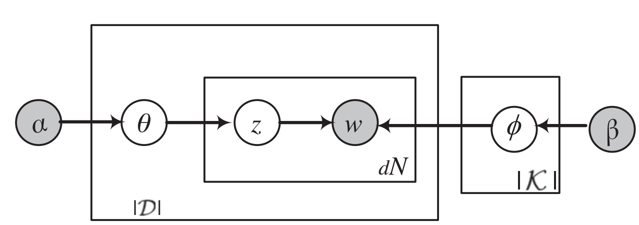

Fig. 1. The graph model of the LDA topic model.

where

$$p(\theta, \mathbf{z}, \mathbf{w}|\alpha, \beta) = p(\theta|\alpha)\prod_{i=1}^{N_d} p(z_i|\theta)p(w_i|z_i, \beta)$$

and

$$p(\mathbf{w}|\alpha, \beta) = \int p(\theta|\alpha)\left(\prod_{i=1}^{N_d}\sum_{z_i} p(z_i|\theta)p(w_i|z_i, \beta)\right)d\theta$$

In multi-aspect sentiment analysis, a social review can be viewed as a document. If all the reviews are directly trained with LDA model, all the topics discussed in these reviews can be got, which are called global topics. However, because LDA model uses the bag-of-words representation of documents, it can explore co-occurrences at the document level, but it cannot preserve the local association between each topic and its sentiment. To address this problem, a sliding window method is proposed as follows.

3.2. Multi-aspect local topic discovery

Definition 1 Multi-aspect sentiment analysis. Assume that a collection of reviews $\mathcal{D}$ contains a set of latent topics $\mathcal{K} = \{\phi_1, \phi_2, \ldots, \phi_{|\mathcal{K}|}\}$, and the set of sentiment expressed in the reviews denoted by a sentiment label set $\mathcal{L} = \{l_1, l_2, \ldots, l_{|\mathcal{L}|}\}$. Given a new review text **d**, the task of multi-aspect sentiment analysis is to discover the topic subset $\mathcal{K}_d$ ($\mathcal{K}_d \subseteq \mathcal{K}$) and associated sentiment label $\mathcal{L}_d$ of the text **d**.

Some researchers have shown that the most important words to determine the topics are the nouns (which also is called name entities), and the words to express sentiment are the adjective, adverbs, and verbs. Because the sentiment words are closely associated with those entities and issues which are evaluated, it is not appropriate to directly segment the topics of the whole review text with trained LDA model. To address this issue, we propose a sliding windows based method. sliding windows are moved from beginning to end in the review text. Each time, only the words in a sliding window are analyzed to discover its local topic and associated sentiment. We assume that a sliding window of the social reviews only presents one topic. In order to distinguish the global topics of the trained LDA model, the topics of sliding windows are called local topics.

One way to discover the local topic would be to consider co-occurrences at the sliding windows, i.e., apply LDA model to individual sliding windows. But in this way there is no a sufficient co-occurrence domain. It is known that LDA behaves badly when applied to short documents [9]. Due to its high dimensional yet extremely sparse representations, clustering short sliding windows directly based on the bag-of-words representation can be very ineffective [23,24]. For this reason, we compute the average KL-divergence between word distribution of a specific global topic and the word distribution of the sliding window context, and then select the topic label with the minimum distance as the local topic label of the sliding window. Afterward, the sentiment label of the associated sentiment is calculated by HowNet lexicon. Finally, we summarized the same local topics and associated sentiments to get the multi-aspect sentiments of the whole review.

Definition 2 Sliding window. Given a word sequence of a review **d** = $\{w_{d1}, w_{d2}, \ldots, w_{dN}\}$, a sliding window $\mathcal{C} = \{w_1, w_2, \ldots, w_{|\mathcal{W}|}\}$ is a sub-sequence of **d**, that is $\mathcal{C} \subset \mathbf{d}$. Any two sliding windows $\mathcal{W}_i$ and $\mathcal{W}_j$ of the same review satisfy $\mathcal{W}_i \cap \mathcal{W}_j = \emptyset$, and all the sliding windows of the review satisfy $\cup_i \mathcal{W}_i = \mathbf{d}$.

In LDA model, a latent topic ϕ_k is a multinomial distribution about the word vocabulary $\mathcal{V}$. For simplicity, we denote the distribution of the topic ϕ_k as $P_k(w_i)$. After LDA model has been trained, the distributions of all the $|\mathcal{K}|$ topics also are determined. A sliding window $\mathcal{W}$ also can be viewed as a multinomial distribution of the vocabulary $\mathcal{V}$. We denoted it as $Q(w_i)$. Then the Kullback–Leibler (KL) divergence can be used to calculate the dissimilarity between the probability distributions of $P_k(w_i)$ and $Q(w_i)$. The KL divergence between the sliding window $\mathcal{W}$ and the latent topic ϕ_k can be calculated by the following equation:

$$KL(Q||P_k) = \sum_{i}^{|\mathcal{V}|} Q(w_i)\log Q(w_i)/P_k(w_i) \quad (2)$$

Because the KL divergence is not symmetrical, in this paper the average KL divergence is used between them as in the following equation:

$$KL(Q, P_k) = (KL(Q||P_k) + KL(P_k||Q))/2 \quad (3)$$

The probability value of a word w_i in the sliding window $\mathcal{W}$ is calculated by the word's occurrence frequency, that is $Q(w_i) = count(w_i)/|\mathcal{W}|$, where $count(w_i)$ is the occurrence number of the word w_i in $\mathcal{W}$, and $|\mathcal{W}|$ is the total word number of $\mathcal{W}$.

Definition 3 Topic discovery of the sliding window. Given a sliding window $\mathcal{W} = \{w_1, w_2, \ldots, w_{|\mathcal{W}|}\}$, a trained LDA topic model which contains the topic set $\mathcal{K}$, and the distribution of topic ϕ_k is $P_k(w_i)$,the topic of the sliding window $\mathcal{W}$ is the one satisfying $\arg\min_{\mathcal{K}} KL(Q||P_k)$.

During the real computation process, where the local topics of some sliding windows are not obvious, in order to discover the local topic more accurately, two threshold parameters ε_1 and ε_2 are used to improve the local topic discovery. When the conditions of $\arg\min_{\mathcal{K}} KL(Q, P_k) < \varepsilon_1$ and $\arg\min_{\mathcal{K}} \frac{2|KL(Q,P_k) - KL(Q,P_j)|}{KL(Q,P_k) + KL(Q,P_j)} \geqslant \varepsilon_2$ $(j = 1, 2, \ldots, |\mathcal{K}|,\ j \neq k)$ both are satisfied, the corresponding topic is selected as the local topic of the sliding window. Otherwise, the local topic of the sliding window is labeled as an extra "other topic".

Definition 4 Sentiment analysis of Sliding windows. Given a sliding window $\mathcal{W} = \{w_1, w_2, \ldots, w_{|\mathcal{W}|}\}$, the sentiment analysis of the sliding window is to identify the sentiment polarity of the sliding window according to some classification function $f_{\mathcal{W}2\mathcal{L}} : \mathcal{W} \to \mathcal{L}$.

The sentiment analysis of sliding windows is calculated based on HowNet lexicon in this paper. We describe it in the following section.

4. Sentiment analysis of sliding windows based on HowNet lexicon

Given a sliding window $\mathcal{W} = \{w_1, w_2, \ldots, w_{|\mathcal{W}|}\}$, a HowNet lexicon based method are proposed to identify its sentiment polarity. The basic idea is as follows: the orientation of the sentiment words in the sliding window is calculated through some positive and negative benchmark words, and then the orientation of the sliding window is summarized from its sentiment words. So the key problem is how to determine the sentiment polarity of a sentiment word. We solve this problem by applying HowNet lexicon.

4.1. Introduction to HowNet lexicon

HowNet lexicon [15] is a general knowledge base of Chinese and English words. There are two basic terminologies in HowNet, which are "concept" and "sememe". Concept is a description of the semantic of words. Every word can express one concept or a few concepts. Sememe denotes the minimum semantic unit to describe the concept. Each concept in HowNet lexicon is described by

several sememes, and the sememes are organized to form a tree structure according the relation of hyponyms and hypernym. The fundamental function of HowNet lexicon is to uncover the associative relationship among the concepts and relations between the attributes of concepts.

The similarity between two words in HowNet lexicon is calculated by the similarity between their concepts. Furthermore, the similarity of two concepts is calculated by the similarity of their sememes. Given two sememes p_i and p_j, their similarity is calculated by their path distance in their sememe tree [43] as shown in the following equation:

$$Sim_p(p_i, p_j) = \frac{\alpha}{d_{ij} + \alpha} \tag{4}$$

where $d_{ij} > 0$ is a positive integer which denotes the path length of p_i and p_{jo} in sememes tree, and α is an adjustable parameter.

The similarity $Sim(S_i, S_j)$ of two concepts S_i and S_j is determined by their maximal sememe similarity. Given two Chinese words w_i and w_j, if w_i contains N concepts denoted by $S_{i1}, S_{i2}, \ldots, S_{iN}$, and w_j contains M concepts denoted by $S_{j1}, S_{j2}, \ldots, S_{jM}$, then the similarity between w_i and w_j equals to the maximum similarity of all the concepts of the words, as shown in Eq. (5).

$$Sim_w(w_i, w_j) = \max_{r,s} Sim(S_{ir}, S_{js}) \quad \text{where} r = 1, 2, \ldots, N, \; s = 1, 2, \ldots, M \tag{5}$$

4.2. How to calculate the sentiment orientation

Turney et al. [32,34] have given a method for inferring the sentiment orientation of a word from its statistical association with a set of positive and negative paradigm words. Zhu et al. [43] applied Turney's idea to calculate the sentiment orientation of Chinese words based on HowNet lexicon. To calculate the sentiment polarity of a sentiment word, at first we choose m pairs of words with strong positive and negative polarity as benchmark words. In the benchmark words, the polarities of positive words are set as +1 and that of negative words are set as −1. Afterward, we calculate the similarity between the sentiment words with those benchmark words and HowNet lexicon. The similarity value is used to measure the word's sentiment orientation.

Given a sentiment word w, its sentiment orientation $O(w)$ will be calculated as Eq. (6), where kp_i denotes the positive benchmark words with the polarity +1 and kq_i denotes the negative benchmark words with the polarity −1.

$$O(w) = \sum_{i=1}^{m} (Sim_w(kp_i, w) - Sim_w(kq_i, w)) \tag{6}$$

Obviously, if $O(w) > 0$, the sentiment polarity of w is positive; if $O(w) < 0$, that means the sentiment polarity of w is negative; otherwise $O(w) = 0$ implies a neutral sentiment polarity of w.

It is worth noting that in Eq. (6) it assumes that all of the benchmark words have same weights to their word similarity. For simplicity, we called Eq. (6) a not weighted method (NWM). More reasonably, the benchmark words of much more similarity should have larger weights. Based on this idea, we propose the benchmark weighted method (BWM) shown in the following equation:

$$O(w) = \sum_{i=1}^{m} (wp_i \cdot Sim_w(kp_i, w) - wq_i \cdot Sim_w(kq_i, w)) \tag{7}$$

where $wp_i = Sim_w(kp_i, w)/\sum Sim_w(kp_i, w)$, $wq_i = Sim_w(kq_i, w)/\sum Sim_w(kq_i, w)$.

Based on (7), the sentiment orientation of a sliding window $O(\mathcal{W})$ can be calculated by the average sum of those orientation of its sentiment words, that is $O(\mathcal{W}) = \sum_{|\mathcal{W}|} O(w_i)/|\mathcal{W}|$. We call it the direct average sum method (DASM).

Furthermore, because the sentiment words in a slide window could have different importance, we can consider the weighted average sum sentiment orientation according to the sentiment words' weights. Then the orientation of a sliding window $\mathcal{W}$ is:

$$O(\mathcal{W}) = \sum_{|\mathcal{W}|} wt_i \cdot O(w_i)/|\mathcal{W}| \tag{8}$$

where wt is the weight of w. In this paper, TFIDF weight of sentiment word is used. We call it the weighted average sum method (WASM).

Table 1
The sentiment tag distribution of the 2000-SINA-Blog dataset.

Sentiment tag	Positive	Neutral	Negative
The number of blogs	843	365	792

5. Experimental setting

In this section, we present the experimental setting of sentiment orientation calculation and topic discovery based on the 2000-SINA-Blog dataset and the 300-SINA-Blog dataset. The 2000-SINA-Blog dataset consists of 2000 blog text with their overall sentiment polarities labeled either positive, neutral or negative, which is used to evaluate the global topic discovery and sentiment orientation calculation in document level. The 300-SINA-Blog dataset consists of 300 blog text with sentiment polarity labels in their paragraphs. It is used to evaluate the local topic discovery and sentiment orientation calculation in paragraph level.

5.1. 2000-SINA-Blog dataset

Because there are no authoritative Chinese social review datasets, we use the web crawler software MetaSeeker[1] to grab Blog pages from the SINA Blog[2] about the social event "real-name system of train tickets (火车票实名制)". The title, text content, and replies of each blog are extracted. Some replies with too short text and insignificant contents are discarded. Then three experienced persons are invited to label each blog text's sentiment tags manually. As it is difficult to give an exact sentiment value, the sentiment tags only include "positive", "neutral", and "negative". The blogs which get consistent sentiment tags are reserved, and those not getting consistent sentiment tags are discarded. At last, we get a social review dataset containing 2000 SINA Blog texts. For simplicity, we call it 2000-SINA-Blog dataset. The sentiment tag distribution of the 2000-SINA-Blog dataset is listed in Table 1, where 843 blogs are labeled as positive, 365 blogs are labeled as neutral, and 792 blogs are labeled as negative.

5.2. 300-SINA-Blog dataset

Labeling local topics and their sentiment is very tedious and time-consuming. Moreover, some topics are not easy to determine. In order to evaluate the experiment results, we select 300 social reviews from the 2000 SINA Blog text. The topics of the 300 blog texts are summarized to 5 aspects: topic 1 is "the cost of the real-name system of train tickets (火车实名制投入成本)"; topic 2 is "the real-name system of train tickets will lead to information loss (火车实名制导致信息丢失)"; topic 3 is "the real-name system of train tickets will combat scalpers (火车实名制打击黄牛党)"; topic 4 is "the experiments of the real-name system of train tickets (火车

[1] http://www.gooseeker.com/cn/node/product/front.
[2] http://blog.sina.com.cn/.

Table 2
The local topic distribution of the 300-SINA-Blog dataset.

Topic category	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Others
The number of paragraphs	213	365	670	408	327	625
The number of paragraphs with sentiment tags	201	316	621	399	272	126

Topic 1: the cost of the real-name system of train tickets (火车实名制投入成本).
Topic 2: the real-name system of train tickets will lead to information loss (火车实名制导致信息丢失).
Topic 3: the real-name system of train tickets will combat scalpers (火车实名制打击黄牛党).
Topic 4: the experiments of the real-name system of train tickets (火车实名制试点).
Topic 5: the certificates required to buy tickets and get into train station in the real-name system of train tickets (火车实名制进站、买票所需要证件).

Table 3
Forty pairs of sentiment benchmark words.

Benchmark words with sentiment value +1				
健康/healthy	友善/friendly	美丽/beautiful	保险/assure	卫生/cleanly
天使/angel	精英/elite	权威/authoritative	优秀/excellent	精选/well-chose
欢喜/joyful	幸福/happiness	容易/easy	文明/civilized	积极/active
著名/notability	完美/perfect	简单/simple	和平/peace	开通/accomplish
真实/real	先进/advanced	便宜/inexpensive	正确/correct	成熟/mature
诚信/faithful	乖巧/clever	精神/vigour	坚定/firm	勤俭/hardworking and thrifty
茂盛/exuberant	安静/peaceful	成绩/good result	雄心/ambition	奖牌/medal
完整/intact	新/new	亮点/lightspot	捷报/news of victory	利润/profit
Benchmark words with sentiment value −1				
事故/accident	黑客/hacker	疯狂/insane	错误/error	不合作/noncooperation
非法/illegality	失败/failure	背后/at the back	麻烦/troublesome	不良/bad
病人/patient	黄色/yellow	色情/erotic	暴力/violence	恶意/malevolence
浪费/waste	漏洞/leak	讨厌/bothersome	自负/overconfident	不安/uneasy
花样/trick	陷阱/trap	敌对/hostility	失误/mistake	流氓/immoral
虚假/sham	变态/abnormal	脆弱/weak	不合格/unqualified	愚/foolish
谣言/rumour	淫秽/bawdry	嘈杂/noisy	残/incomplete	恶势力/vicious power
缺少/lack	脏/dirty	陈旧/obsolete	丑陋/ugly	毒/poison

实名制试点)''; and topic 5 is ''the certificates required to buy tickets and get into train station in the real-name system of train tickets (火车实名制进站、买票所需要证件)''. The ''others'' topic is about something that are not included in the above five topics.

Analyzing the 300 social reviews, the total number of the paragraphs is 2608. The paragraph number of the main body is 1842, and the paragraph number of the replies is 767. Each reply is viewed as a paragraph. We use all the reviews to train a LDA model for discovering the global topics. And then identify the local topics with the trained LDA model. Those paragraphs of which local topics have been identified correctly are labeled sentiment tags manually. Three experienced persons are invited to analyze each paragraph of the 300 social reviews, and label the local topic and sentiment tags of each paragraph. The sentiment tags include ''positive'', ''neutral'', and ''negative''. The local topics and sentiment distribution are listed as Table 2.

5.3. Preprocessing and evaluation criteria

The preprocessing procedures of social reviews include Chinese word segment, part of speech tagging, building the vector space model, etc. In this paper, ICTCLAS[3] is used to implement Chinese word segment and part of speech tagging. And the person name, location name, organization name, words denoting time, numeral words, measure words, pronoun, punctuation, etc. are filtered out before calculating the sentiment orientation.

The accuracy of local topic discovery is defined as:

$$accu = \frac{\sum_{k=1}^{|K|} tn_k}{TN - tn_{other}}$$

where tn_k denotes the number of paragraphs which have been correctly identified by the topic k, TN denotes the total number of paragraphs which have been tested, tn_{other} is the number of paragraphs which belong to the ''other'' topic.

Table 4
Test accuracies (in%) in sentiment orientation judgment on 2000 pair oppositive sentiment words from HowNet lexicon.

Algorithms	Commendatory terms	Derogatory term	Average accuracy
NWM	87.5	99.65	93.58
BWM	95.45	98.35	96.90

6. Experimental results

In this section, we present and discuss the experimental results of both sentiment orientation calculation and topic identification.

6.1. Sentiment orientation calculation to sentiment words

To compare the performance of NWM and BWM, forty pairs of benchmark sentiment words are selected according to the Hits value returned by Google search engine. The repeated words are merged out. The forty pairs of benchmark sentiment words in our experiments are listed in Table 3.

We select 2000 pairs of appositive words from HowNet lexicon,[4] which have been respectively labeled ''good'' and ''bad'', acting as the test dataset. The ten-fold cross validation experiments are carried out separately with NWM and BWM. The experiment results are shown in Table 4. According Table 4, the BWM is better than the original equation in identifying the commendatory terms. The identification accuracy is improved from 87% to 95.45%. And to derogatory term, the identification accuracy is slightly decreased from 99.65% to

[3] http://ictclas.org/.

[4] http://www.keenage.com.

Table 5
The sentiment orientation values of some feature terms.

Positive sentiment		Neutral sentiment		Negative sentiment	
Words	Values	Words	Values	Words	Values
信心/confidence	0.62566	代表性/representative	0	劳民伤财/waste	−0.86823
不错/right	0.36857	正是/no other than	0	侵犯/violate	−0.84344
有用/useful	0.35714	铁路局/railway bureau	0	片面/single-faceted	−0.50283
公平/fair	0.30944	推开/push	0	耗费/cost	−0.44112
民意/public opinion	0.28360	考量/judge	0	负担/load	−0.41284
保障/guarantee	0.26141	试想/imagine	0	官僚/bureaucrat	−0.35435
立场/standpoint	0.24576	票务/ticket business	0	欠佳/below the average	−0.25959
利益/benefit	0.18042	公里数/kilometer of travel	0	死板/inflexible	−0.24290
满足/satisfied	0.15912	还有/in addition	0	借口/excuse	−0.23573
刷新/refresh	0.12009	看似/like	0	不便/inconvenient	−0.18069

Table 6
Test accuracies (in %) in document level with different u value.

The value of parameter u	0.005	0.01	0.015	0.02	0.025
BWM + DASM	54.33	65.96	73.28	70.03	64.54
BWM + WASM	55.74	68.65	75.89	73.21	66.87

98.35%. However, as the average accuracy is improved from 93.58% to 96.90% by BWM, BWM is used in the following experiments.

Analyzing the test words, we find most of them have strong sentiment orientation. NWM and BWM can determine their sentiment orientation correctly. But there are a few words which have not obvious orientation. These words have different similarity values with two kind benchmark words. BWM can give larger weights to the more similar benchmark words. It enlarges the sentiment difference, and improves the sentiment orientation classification. NWM gives the same weight to all the benchmark words, sometimes it cannot correctly identify the orientations of obscure sentiment words.

6.2. Sentiment analysis in document level

Given a social review d, when the sliding window satisfies $\mathcal{W} = d$, we can analyze the sentiment orientation in document level. At first, BWM is used to calculate the feature terms' sentiment orientation. The orientation values of some feature terms are listed in Table 5. From Table 5, we can see that the orientation values of feature terms can be used to identify the sentiment words and non-sentiment words. If a feature term has larger absolute value of sentiment orientation, it means that this feature term contributes larger weight to the sentiment orientation of the sliding window.

After the feature terms' orientation values have been calculated, DASM and WASM are used separately to calculate the review's sentiment orientation. A sentiment threshold parameter u is introduced to define the sentiment category. If $-1 \leqslant O(d) < -u$, the sentiment orientation of d is negative; if $-u \leqslant O(d) \leqslant u$ then neutral, and if $u < O(d) \leqslant +1$ then positive. Ten-fold cross validation experiments are carried out with different threshold values. The accuracy results of orientation identification in document level with different u value are shown in Table 6.

From Table 6, we can see that the parameter u influences the performance of the sentiment classification. When u is set as 0.015, Both BWM + DASM and BWM + WASM get the best sentiment classification accuracy. Furthermore, in Table 6, all the test accuracies of WASM are better than that of DASM. The experiment results show that summarizing the review's sentiment orientation according the sentiment words' weights can improve the review's sentiment orientation calculation. However, the best sentiment classification accuracy of BWM + DASM is 73.28%, and BWM + WASM is 75.89%. Obviously these best accuracies are not satisfying. The primary reason of the poor performance lies in the coarse-grained analysis. Because the multi-aspect topics in a social review are varied, they are not always consistent in sentiment orientation and the sentiment orientation in document level is not determinate. It is necessary to do fine grain and multi-aspect sentiment analysis.

6.3. Multi-aspect sentiment analysis

Multi-aspect sentiment analysis contains two subtasks: one is aspect extraction, and the other is sentiment orientation calculation of aspects. The experimental results of local topic discovery and sentiment analysis of aspects will be given in this section.

6.3.1. Global topic discovery with LDA model

In this study, Mark's Matlab Topic Modeling Toolbox[5] is used to discover global topics. We train a LDA model with all the 300 social reviews of 300-SINA-Blog dataset without manual labels. Different topic number $|\mathcal{K}|$ of LDA model are tested. There are works show that the average similarity of the topic probabilities is a good measure to optimize the topic parameter of the LDA model. For example, Cao et al [3] argued that the optimal LDA model can be obtained when the average similarity of the topic structure is minimized. So the sum of average KL distance between each topic is computed to determine the appropriate topic number $|\mathcal{K}|$ in our experiments. The experiment results are listed in Table 7.

According to Table 7, the sum of average KL distance of each topic gets the maximum values when the topic number is 5 or 6. The 15 words having the maximum probability values in each topic are list in Table 8. These topic words can represent the global topics reasonably.

6.3.2. Local topic discovery with sliding windows

We test three types of sliding window, one is based on sentences, one is base on paragraphs, and the other is dynamic. The sentence based method selects each sentence as a sliding window, the paragraph based method selects each paragraph as a sliding window, and the dynamic method selects several sentences by the optimized KL distance. The experiment results in a small dataset show that the average accuracies of the sentence based method and the dynamic method both are less than 50%, and the average accuracy of the paragraph based method achieves 72.97%.

Analyzing the experiment results, we find the reason of the low accuracy of the sentence based method is that the sentence contains too few words, which leads the computed KL divergence value is poor to is criminate different topics. And the reason of the low accuracy of the dynamic span method is that the manual label

[5] http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm.

Table 7
The sum of average KL distance results with different topic numbers.

Topic number $\|\mathcal{K}\|$	3	4	5	6	7	8	9	10
The summarized average KL distance	334.80	419.09	442.24	443.64	390.60	348.24	315.83	295.35

Table 8
The feature words which have the maximization weights of five topics.

	Words	Weights	Words	Weights	Words	Weights
Topic 1	中国/China	0.04523	时间/time	0.02072	实名/real name	0.01226
	铁路/railway	0.04187	北京/Beijing	0.01576	服务/service	0.01109
	可以/not bad	0.02582	火车/train	0.01547	购票/buy ticket	0.01094
	成本/cost	0.02334	验票/ticket check	0.01532	技术/technology	0.01065
	系统/system	0.02145	国家/country	0.01240	票价/ticket price	0.00992
Topic 2	车票/ticket	0.07012	可以/can	0.02047	个人/personal	0.01285
	记者/reporter	0.04046	身份/identity	0.01920	没有/have not	0.01158
	信息/information	0.03586	可能/may	0.01745	火车站/railway station	0.01063
	实名制/real name	0.02729	火车票/ticket	0.01523	深圳/Shenzhen	0.01063
	身份证/identity card	0.02189	号码/number	0.01412	还有/have	0.01063
Topic 3	实名制/real name	0.05679	春运/transport during the Spring Festival	0.01945	部门/department	0.01271
	火车票/ticket	0.05679	倒票/ticket scalping	0.01725	刘志军/Zhijun Liu	0.01245
	铁路/railway	0.04642	一票难求/hard-to-get tickets	0.01660	出行/trip	0.01245
	票贩子/scalper	0.02334	买票/buy ticket	0.01660	制度/rules	0.01232
	实行/implement	0.02308	社会/society	0.01413	应该/should	0.01232
Topic 4	实名制/real name	0.15480	成都/Chengdu	0.02421	买票/buy ticket	0.01593
	火车票/ticket	0.09258	试点/experimental unit	0.01932	实行/implement	0.01493
	春运/transport during the Spring Festival	0.05006	广铁/Guangzhou Railway Group	0.01819	全国/nationwide	0.01455
	广州/Guangzhou	0.04868	火车站/railway station	0.01794	需要/need	0.01305
	旅客/traveller	0.02798	试行/try out	0.01706	集团/group	0.01167
Topic 5	身份证/ID Card	0.07026	成都/Chengdu	0.02218	乘车/get on the train	0.01437
	旅客/traveller	0.06109	身份/identity	0.01958	工作/work	0.01413
	车站/rail station	0.03953	证件/certificate	0.01809	铁路/railway	0.01400
	车票/ticket	0.03123	信息/information	0.01648	部门/department	0.01400
	进站/get in sation	0.02577	证明/certified	0.01450	购买/purchase	0.01351

Table 9
The results of local topic discovery with different threshold value ε_1 and ε_2.

ε_1	0.03	0.04	0.05	0.06	0.07
ε_2	0.1	0.15	0.2	0.25	0.3
Accuracy (%)	79.43	85.78	91.23	89.61	74.03

work is very difficult. The three persons invited to label the sliding windows frequently have different opinions about a sentence's topic. There are many difference in the labeled dataset.

Due to the page limit, we only evaluate the paragraph based method in this paper. The above trained LDA model with global topics is used to discover each paragraph's local topic of 300-SINA-Blog dataset. The experiment results of different parameters ε_1 and ε_2 are shown in Table 9. From Table 9, we can see that when $\varepsilon_1 = 0.05$ and $\varepsilon_2 = 0.2$, the best accuracy of local topic discovery is 91.23%.

The identification results of each local topic are listed in Table 10, where the second row provides the actual identification results with the trained LDA model, and the third row provides the correct identification result of each topic. Among the 2608 paragraphs of 300-SINA-Blog dataset, the total number of paragraphs whose local topics have been correctly identified is 1809. But because the "others" topic has unclear topic, if we exclude the paragraphs belong to the "other" topic, the average accuracy of local topic identification on topic 1 to topic 5 can achieve 91.23%. This experimental result shows that the trained LDA model is good to determine the paragraphs' local topics.

6.3.3. Multi-aspect sentiment analysis

The above 1809 paragraphs whose local topics have been correctly identified are used to evaluate the paragraphs' sentiment orientation. We calculate these paragraphs' sentiment orientations with BWM + DASM and BWM + WASM separately. With BWM + DASM, 593 paragraphs are identified to be positive sentiment orientation, 427 paragraphs are neutral, and 789 paragraphs are negative. There are total 1613 paragraphs correctly identified, and the average accuracy of sentiment analysis of all topics is 89.17%. With BWM + WASM, 569 paragraphs are identified to be positive, 443 paragraphs are neutral, and 797 paragraphs are negative. There are total 1667 paragraphs which have been identified correctly, and the average accuracy of sentiment analysis of all topics is 92.15%. The sentiment analysis results of each topic's paragraphs are listed in Table 11.

According Table 11, we find the accuracy of the sentiment analysis in each topic is higher than the accuracy in the whole document level. Furthermore, the reviewer's sentiment expressed in each topics can be known clearly.

6.3.4. Comparing with other methods

We also compare the proposed MSA-COSR method with some supervised methods and other topic model methods. Taking the 1983 paragraphs of the topic 1-5 in Table 7 as experiment dataset, some supervised method such as Naïve Bayesian (NB), k-nearest neighbors (KNN) and support vector machine (SVM) are selected to classify the paragraphs' sentiment orientation. The algorithm implementations of NB, KNN, and SVM come from the Weka toolkit.[6] The optimization parameters of each classification method are determined by the Weka toolkit itself. Three-fold cross validation is executed in the experiments. The best classification accuracies of each method are selected for comparison. All results are listed in

[6] http://www.cs.waikato.ac.nz/ml/weka/.

Table 10
Topic identifying results with LDA model.

Topic category	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	others
The identification results by LDA model	213	365	670	408	327	625
The correct identification results	201	316	621	399	272	126

Table 11
The sentiment analysis results of different topics.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
The paragraphs' number used to evaluate	201	316	621	399	272
The paragraphs' number identified correctly with BWM + DASM	179	287	563	326	258
The paragraphs' number identified correctly with BWM + WASM	183	296	581	347	260
Identification accuracies (%) with BWM + DASM	89.05	90.82	90.66	81.70	94.85
Identification accuracies (%) with BWM + WASM	91.04	93.67	93.55	86.08	95.59

Table 12
Topic identifying results of various algorithms in 300-SINA-Blog dataset.

Algorithm	NB	KNN	SVM	LDA	JST	MSA-COSR
Accuracy (%)	77.36	86.74	83.21	78.55.	78.96	92.15

Table 12.

We also compare MSA-COSR with LDA model and JST model. Because both LDA and JST are document level topic models, each paragraph of the above 1983 paragraphs is viewed as a document to train LDA and JST model. It is the same as MSA-COSR, Matlab Topic Modeling Toolbox is used to implement LDA, and the topic number is set as 6. The topic label of each paragraph is determined by the maximum probabilistic topic of the LDA. And Phan's GibbsLDA++ package[7] is modified to implement JST model. The topic number is set as 6, and the sentiment labels are set as "positive", "neutral" and "negative". The topic identification accuracies of LDA and JST are also listed in Table 12. The overall sentiment identification accuracy of JST without prior information is only 58.72%. From Table 12, it is obviously that MSA-COSR is better than the other comparative methods.

7. Conclusion and future works

In this paper, we proposed a multi-aspect sentiment analysis of Chinese social reviews based on LDA topic model and HowNet lexicon. Our method adopts trained LDA model to identify the local topic of the sliding windows, and then analyze the associated sentiment of these sliding windows with HowNet lexicon. The experiment results show that our method not only obtains good performance in topic identification, but also has good performance in the sentiment analysis. The best accuracy of the topic identification is 91.23%, and the best average accuracy of the sentiment judgment with BWM+WASM is 92.15%. Our method can identify multi-aspect topic and their sentiment orientation in Chinese social review. It is useful to handle the sentiment analysis task, and help us to analyze the sentiment orientation more accurately and deeply.

Despite the success of the proposed method, it is difficult to select a suitable topic number to train the LDA model. In the future works, we will consider some available methods such as hierarchical Dirichlet process to resolve this problem. In addition, to compare our method with the topic and sentiment model such as TSM, JST in other datasets, and to analyze the topic evolution over time also are future works .

7 http://gibbslda.sourceforge.net/.

Acknowledgments

The authors would like to thank anonymous reviewers for their constructive reviews and good advices, and thank Zhexuan Zhu and Wenfeng Du for helpful discussion. This research was supported by National Natural Science Foundation of China (61003271, 61001185), Science and Technology Foundation of Shenzhen City (JC201005280463A, JC201105160498A).

References

[1] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[2] M. Cannataro, P.H. Guzzi, P. Veltri, Protein-to-protein interactions: technologies, databases, and algorithms, ACM Comput. Surv. 43 (2010) 1–36.
[3] J. Cao, D. Zhang, J. Li, S. Tang, A method of adaptively selecting best LDA model based on density, Chinese J. Comput. 31 (2008) 1780–1787.
[4] W. Du, S. Tan, X. Yun, X. Cheng, A new method to compute semantic orientation, J. Comput. Res. Develop. 46 (2009) 1713–1720.
[5] M. Ganapathibhotla, B. Liu, Mining opinions in comparative sentences, in: Proceedings of the 22nd International Conference on Computational Linguistics, vol. 1, Publishing, Manchester, United Kingdom, 2008, pp. 241–248.
[6] T.L. Griffiths, S. Mark, Finding scientific topics, in: Proceeding of the National Academy of Sciences of the United States of America. Publishing, 2004, pp. 5228–5235.
[7] Y. Hu, R. Lu, X. Li, J. Duan, Y. Chen, Research on language modeling based on sentiment classification of text, J. Comput. Res. Develop. 44 (2007) 1469–1475.
[8] M. Jiang, S. Argamon, Exploiting subjectivity analysis in blogs to improve political leaning categorization, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Publishing, Singapore, Singapore, 2008, pp. 725–726.
[9] O. Jin, N.N. Liu, K. Zhao, Y. Yu, Q. Yang, Transferring topical knowledge from auxiliary long texts for short text clustering, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Publishing, Glasgow, Scotland, UK, 2011, pp. 775–784.
[10] H.D. Kim, C. Zhai, Generating comparative summaries of contradictory opinions in text, in: Proceeding of the 18th ACM Conference on Information and Knowledge Management. Publishing, Hong Kong, China, 2009, pp. 385–394.
[11] F. Li, M. Huang, X. Zhu, Sentiment analysis with global topics and local dependency, in: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10). Publishing, Atlanta, Georgia, USA, 2010, pp. 1371–1376.
[12] X. Liao, D. Cao, B. Fang, H. Xu, X. Cheng, Research on blog opinion retrieval based on probabilistic inference model, J. Comput. Res. Develop. 46 (2009) 1530–1536.
[13] C. Lin, Y. He, Joint sentiment/topic model for sentiment analysis, in: Proceeding of the 18th ACM Conference on Information and Knowledge Management. Publishing, Hong Kong, China, 2009, pp. 375–384.
[14] W.-H. Lin, T. Wilson, J. Wiebe, A. Hauptmann, Which side are you on?: identifying perspectives at the document and sentence levels, in: Proceedings of the Tenth Conference on Natural Language Learning, Publishing, New York City, New York, 2006, pp. 109–116.
[15] Q. Liu, S. Li, Word Semantic Similarity Computation based on Hownet, the 3th Chinese Lexical Semantic Workshop (CLSW2002), Publishing, Taipei, 2002.
[16] R. Malouf, T. Mullen, Taking sides: user classification for informal online political discourse, Internet Res. 18 (2008) 177–190.

[17] Q. Mei, X. Ling, M. Wondra, H. Su, C. Zhai, Topic sentiment mixture: modeling facets and opinions in weblogs, in: Proceedings of the 16th International Conference on World Wide Web. Publishing, Banff, Alberta, Canada, 2007, pp. 171–180.
[18] H.-J. Min, J.C. Park, Toward finer-grained sentiment identification in product reviews through linguistic and ontological analyses, in: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Publishing, Suntec, Singapore, 2009, pp. 169–172.
[19] T. Mullen, R. Malouf, A preliminary investigation into sentiment analysis of informal political discourse, in: AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW). Publishing, 2006, pp. 159–162.
[20] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, A.F. Smeaton, Topic-dependent sentiment analysis of financial blogs, in: Proceeding of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion. Publishing, Hong Kong, China, 2009, pp. 9–16.
[21] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (2008) 1–135.
[22] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10. Publishing, 2002, pp. 79–86.
[23] X.-H. Phan, C.-T. Nguyen, D.-T. Le, L.-M. Nguyen, S. Horiguchi, Q.-T. Ha, A hidden topic-based framework toward building applications with short web documents, IEEE Trans. Knowl. Data Eng. 23 (2011) 961–976.
[24] X.-H. Phan, L.-M. Nguyen, S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections, in: Proceeding of the 17th International Conference on World Wide Web. Publishing, Beijing, China, 2008, pp. 91–100.
[25] V. Satuluri, S. Parthasarathy, Scalable graph clustering using stochastic flows: applications to community discovery, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Publishing, Paris, France, 2009, pp. 737–746.
[26] S. Somasundaran, J. Wiebe, Recognizing stances in ideological on-line debates, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Publishing, Los Angeles, California, 2010, pp. 116–124.
[27] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, Z. Su, Hidden sentiment association in chinese web opinion mining, in: Proceeding of the 17th International Conference on World Wide Web. Publishing, Beijing, China, 2008, pp. 959–968.
[28] H. Tang, S. Tan, X. Cheng, Research on sentiment classification of chinese reviews based on supervised machine learning techniques, J. Chinese Inform. Process. 21 (88–94) (2007) 108.
[29] F. Tao, J. Gao, T. Wang, K. Zhou, Topic oriented sentimental feature selection method for news comments, J. Chinese Inform. Process. 24 (2010) 37–43.
[30] I. Titov, R. McDonald, A Joint Model of Text and Aspect Ratings for Sentiment Summarization, ACL'08. Publishing, 2008a, pp. 308–316.
[31] I. Titov, R. McDonald, Modeling online reviews with multi-grain topic models, in: Proceeding of the 17th International Conference on World Wide Web. Publishing, Beijing, China, 2008b, pp. 111–120.
[32] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Publishing, Philadelphia, Pennsylvania, 2002, pp. 417–424.
[33] P.D. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for, Computational Linguistics, 2002, pp. 417–424.
[34] P.D. Turney, M.L. Littman, Measuring praise and criticism: inference of semantic orientation from association, ACM Trans. Inf. Syst. 21 (2003) 315–346.
[35] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis on review text data: a rating regression approach, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Publishing, Washington, DC, USA, 2010, pp. 783–792.
[36] C. Whitelaw, N. Garg, S. Argamon, Using appraisal groups for sentiment analysis, in: Proceedings of the 14th ACM International Conference on Information and Knowledge Management. Publishing, Bremen, Germany, 2005, pp. 625–631.
[37] Y. Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3, vol. 3. Publishing, Singapore, 2009, pp. 1533–1541.
[38] J. Xu, Y. Ding, X. Wang, Sentiment classification for Chinese news using machine learning methods, J. Chinese Inform. Process. 21 (2007) 95–100.
[39] L. Xu, H. Lin, Z. Yang, Text orientation identification based on semantic comprehension, J. Chinese Inform. Process. 21 (2007) 96–100.
[40] C. Yang, S. Feng, D. Wang, N. Yang, G. Yu, Analysis on web public opinion orientation based on extending sentiment lexicon, J. Chinese Comput. Syst. 31 (2010) 691–695.
[41] Q. Ye, Z. Zhang, R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Syst. Appl. 36 (2009) 6527–6535.
[42] W.X. Zhao, J. Jiang, H. Yan, X. Li, Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Publishing, Cambridge, Massachusetts, 2010, pp. 56–65.
[43] Y. Zhu, J. Min, Y. Zhou, J. Huang, L. Wu, Semantic orientation computing based on HowNet, J. Chinese Inform. Process. 20 (2006) 14–20.